AD_____

Award Number: W81XWH-08-1-0642

TITLE:

Identification of Breast Cancer Serum Biomarkers: A Novel Retroviral Library Screen to Define the Breast Cancer-Soluble Ectodomain Proteome

PRINCIPAL INVESTIGATOR: Dr. Amanda Paulovich, MD, PhD

CONTRACTING ORGANIZATION: Fred Hutchinson Cancer Research Center
Seattle, WA 98109

REPORT DATE: October 2009

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

√ Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-10-2009	Final	9-29-2008 to 9-28-2009

4. TITLE AND SUBTITLE
Identification of Breast Cancer Serum Biomarkers: A Novel Retroviral Library Screen to Define the Breast Cancer-Soluble Ectodomain Proteome

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-08-1-0642

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Biing Y. Lin, Kyle J. Garton, and Amanda G. Paulovich

Email: apaulovi@fhcrc.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Fred Hutchinson Cancer Research Center

Seattle, WA 98109-1024

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Material Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We proposed to perform a proteomic screen of a retroviral-expressed human breast cancer cDNA library in an attempt to identify novel secreted or shed proteins as candidate soluble biomarkers in human breast cancer. Although we encountered difficulties in generating a viable cDNA library, we were able to show that C-terminal HA-epitope tagged proteins can be efficiently expressed in human breast cancer cell lines by retroviral gene transfer. In addition, human breast cancer cells have an active ectodomain sheddase machinery and are capable of releasing large amounts of over-expressed protein. Future studies will be done to perfect our retroviral expression vectors and to subclone pre-made cDNA libraries into suitable expression vectors.

15. SUBJECT TERMS
Retroviral cDNA library, soluble biomarkers, breast cancer, retroviral expression, ectodomain shedding, proteomics.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	21	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

I. Introduction

Cancer has affected an estimated 1.5 million people in 2009, and has caused the deaths of more than a half million. Among all cancer-related cases, nearly 200,000 people have been affected specifically by breast cancers, resulting in more than 40,000 deaths in 2009 (NIH SEER Cancer Statistics Review, http://seer.cancer.gov/csr/1975_2006/index.html). The cancer-specific markers that are currently being used in serological tests to screen patients, such as prostate-specific antigen (PSA), CA 15-3, and CA 27.29, cannot be applied to diagnostic or treatment efficacy in all types of prostate and breast cancers. In addition to current cancer screening, diagnostic biomarkers for early detection, diagnosis, prognosis, and treatment for all types of cancers are needed.

Many characteristics of cancer progression can be revealed by histological abnormalities. Molecular changes occur along with these histological changes, including epigenetic status, transcriptional profile, protein modification and localization within the cell. These molecular changes can be a rich source of cancer-specific biomarker candidates. For example, the secreted protein osteopontin (OPN) was found to be elevated in the plasma of a malignant Her2 breast cancer mouse model (1). The extracellular domain (ECD) of membrane proteins such as the ligands of EGFR (or its Her2 family of receptors) like EGF, transforming growth factor (TGF)-α, and amphiregulin, are shed by disintegrin and metalloproteinase (ADAM) on the cell surface in various cell types. In fact, the increased enzymatic activity of metalloprotease ADAM-10 and the release of ECDs of the Her family of proteins (also known as ErbB) are correlated with the progression of breast cancer (2).

We hypothesize that a proteomic approach can be used to detect changes in protein stability, protein secretion and ectodomain shedding of membrane proteins in a cancerous cell, and these changes can provide potential biomarkers candidates in the serum. We designed a genetic screen of a retroviral breast cancer cDNA library, including a custom vector containing an expressible HA-tagged protein at the N'- or C'-terminal end. Using antibody specific to HA, we can harvest and concentrate proteins from both cell lysates and from culture medium. Our hypothesis was that we could isolate tagged proteins from our library that were shed or secreted from breast cancer cells, thus providing potential serum-soluble biomarkers. However, we failed to create a tagged cDNA library. Our alternative approach is to subclone human breast cancer-associated and kinase libraries into our retroviral vector to transduce breast cancer cell lines. To identify the tagged protein, the immunoprecipitates were digested with trypsin into peptides for analysis with liquid chromatography-mass spectrometry (LC-MS). This experimental scheme is illustrated in Figure 1.

Flow Chart of Project

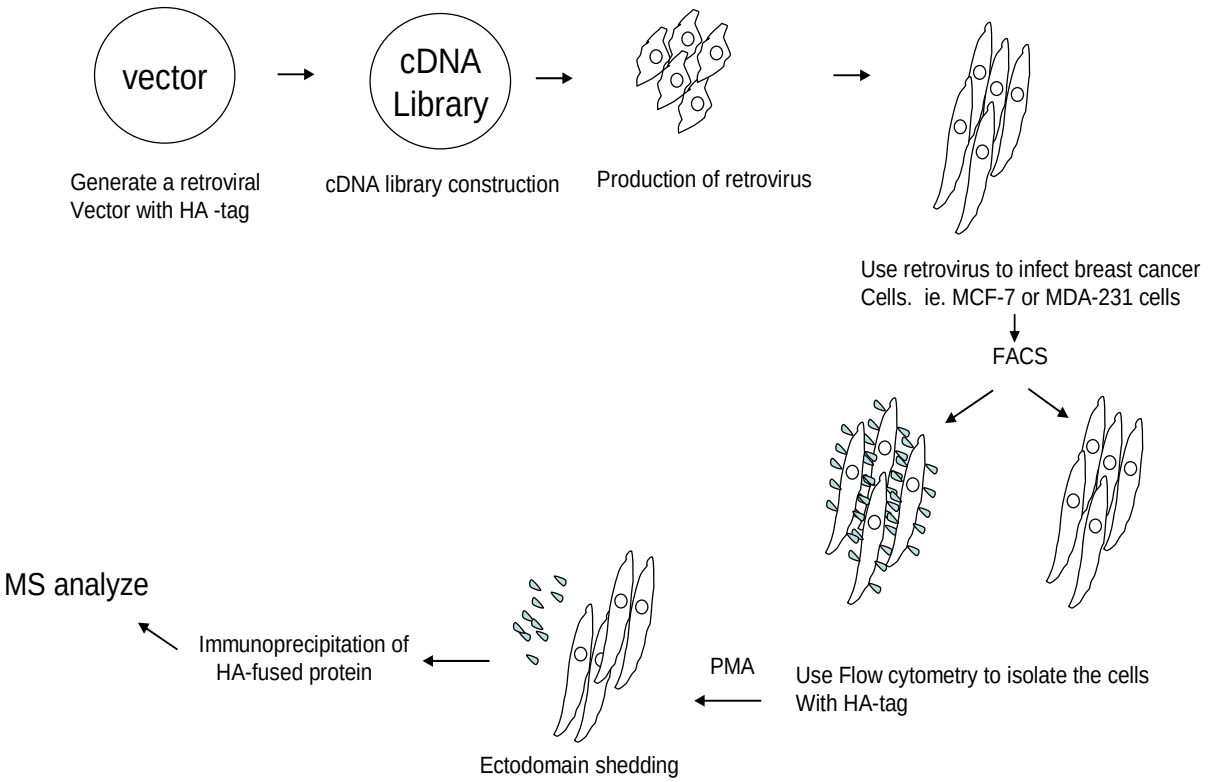


Figure 1: Flow chart of the project using a retroviral vector to stably express breast cancer associated proteins in cultured cells followed by mass spectrometry analysis.

II. Body

PART A

Stable and efficient transgene expression with IRES-based retroviral vectors

An important component of the proposed research project was the ability to efficiently and stably express a variety of transgenes (cDNAs) in multiple different cell lines including human breast cancer cells that can be used in mouse xenograft models. We proposed to synthesize and test retroviral vectors in which individual genes could be expressed with epitope tags at either the N- or C-terminus. The pBM-XHA-IRES-EGFP vector was synthesized by adaptation of the pBM-IRES-EGFP retroviral vector (Gary Nolan, Stanford University) using standard molecular techniques. The pBM-XHA-IRES-EGFP vector is a single transcript expression vector in which two genes are expressed from a single retroviral RNA within transduced cells. Individual genes cloned into the first cistron are expressed with a C-terminal hemaglutinin (HA)

epitope-tag. Incorporation of asymmetric SfiI-restriction sites was also used for efficient directional cloning of amplified cDNAs. Expression of the second cistron reporter genes (Green Fluorescence Protein, EGFP) or other selectable markers is strongly linked to expression of genes cloned into Cistron 1. To test the stability of gene expression achieved with the pBM-XHA-IRES-EGFP vector, a test gene ADAM15 was cloned into the Cistron 1 position (Figure 2A). ADAM15 was amplified by PCR and 5' and 3' SfiI sites were incorporated into the resulting amplification product by incorporation into the upstream and downstream primers (not shown). The resulting PCR product was sequenced and the resulting product digested with SfiI and cloned into SfiI-linearized pBM-XHA-IRES-EGFP. In general using this strategy, greater than 80% of resulting clones contained the proper insert with the proper orientation confirming the utility of asymmetric SfiI sites for directional cloning (not shown). After preparation of high-titer retrovirus, NIH3T3 cells were infected with either the empty retroviral vector or with virus expressing ADAM15. Using flow cytometry, greater than 60% of the cells were infected as evidenced by expression of the EGFP reporter transgene. To measure the stability of transgene expression over time in cultured cells, we used flow cytometry and cell sorting to isolate cells in the highest quartile of EGFP expression (high) and cells in the lowest quartile of expression (low) and subsequently cultured the cells for multiple passages in vitro (Figure 2B). Repeat analysis by flow cytometry showed that cells maintained their relative expression level of EGFP with no significant loss of expression over time. We then determined whether expression of different levels of the EGFP reporter gene were correlated with different expression levels of the ADAM15 gene that was present in the Cistron 1 position. Cells expressing higher levels of EGFP showed higher levels of ADAM15 expression as determined by Western blotting for the HA-epitope tag (Figure 2C). In addition, these relative expression levels for the ADAM15 gene were maintained over multiple passages of in vitro culture with no appreciable loss of expression. Together these experiments showed that the pBM-XHA-IRES-EGFP vector leads to efficient and stable transgene expression that can be readily detected through the HA-epitope tag with stable expression over multiple passages in culture and no apparent loss of expression.

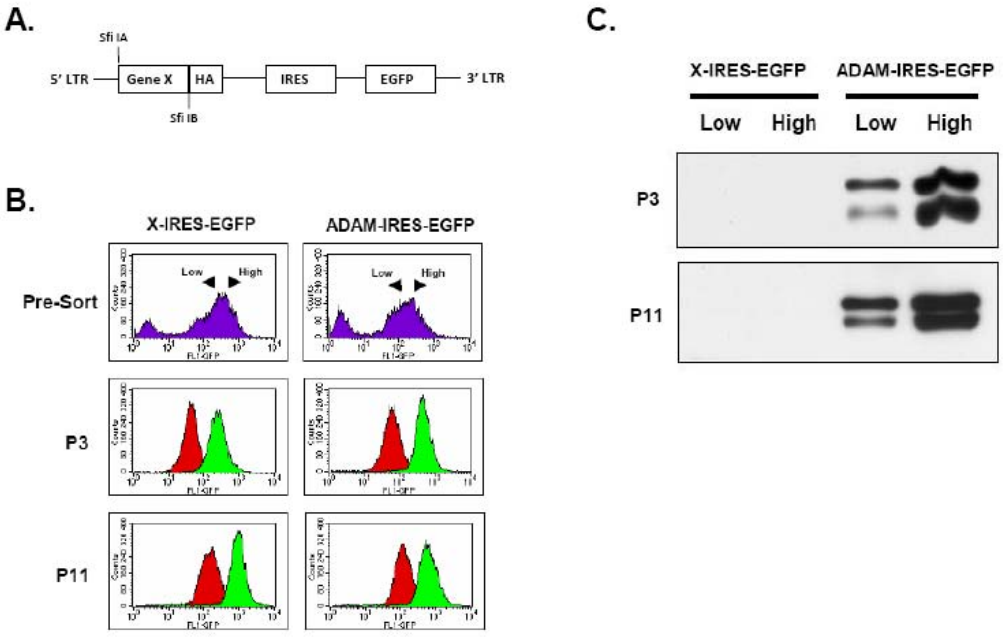


Figure 2: **A.** pBM-XHA-IRES-EGFP retroviral vector structure and design. **B.** Stability of retroviral gene expression. NIH3T3 cells were infected with a retrovirus containing an HA-epitope tagged version of human ADAM15 or an empty vector control. Cells expressing low vs. high levels of the EGFP reporter gene were then isolated and expanded in culture for 3 and 11 passages and expression of the EGFP reporter gene was determined by flow cytometry. **C.** Expression of HA-tagged ADAM15 was determine by Western blotting.

Retroviral Expression and Selection in Human Breast Cancer Cell Lines

To determine the efficiency of retroviral gene expression in human breast cancer cell lines, a retrovirus pBM-EGFP-IRES-PURO was synthesized such that infected cells would express the EGFP gene from the first cistron position and the puromycin resistance gene from the second cistron position. Using either MCF-7 or MDA-231 cells, infection without selection led to relatively low transduction efficiencies of 22% and 16% respectively (Figure 3). However, after selection with puromycin for a single passage, populations of cells with very high transduction rates could be achieved that approached 100%. Therefore, the pBM-based retroviral vectors, coupled with puromycin selection leads to efficient transduction of the human breast cancer cell lines, MCF-7 and MDA-231 (Figure 3).

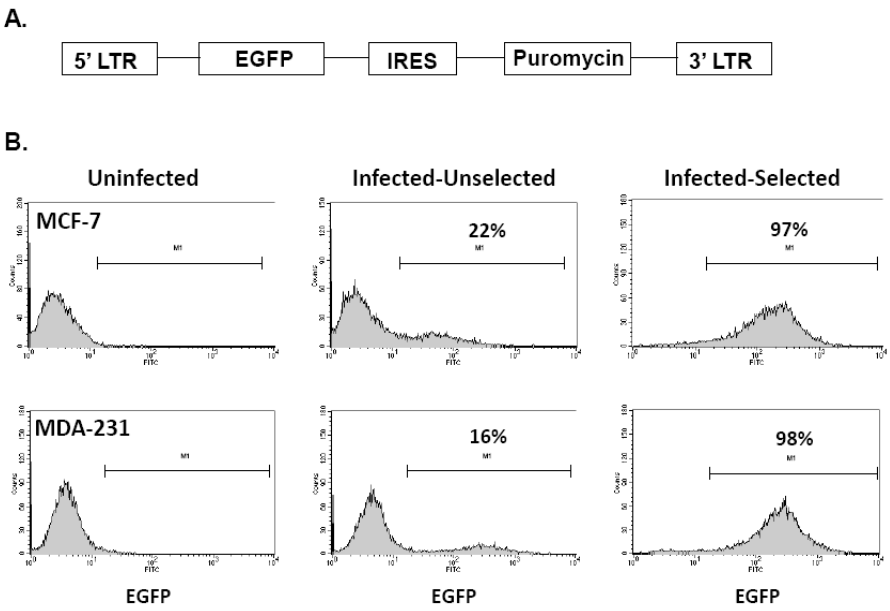


Figure 3: Retroviral expression in human breast cancer cell lines. The human breast cancer cell lines MCF-7 and MDA-231 were infected with retrovirus derived from the pBM-EGFP-IRES-PURO retroviral vectors and infection efficiencies were determined by flow cytometric determination of the percentage of cells expressing the first cistron EGFP. Infection efficiencies were determined before (unselected) and after (selected) selection with puromycin for a single passage in culture.

Human Breast Cancer Cells Contain an Active Ectodomain Sheddase Machinery

We have proposed that soluble serum biomarkers of breast cancer may be generated by a number of mechanisms including shedding of proteins from the cell surface through metalloproteinase mediated shedding, termed ectodomain shedding. L-selectin is a prototype substrate of a cell surface protein in which cell surface levels are actively regulated through ectodomain shedding by the matrix metalloproteinase, ADAM17 (or TACE). We therefore wanted to confirm that human breast cancer cells contain an active ectodomain sheddase machinery and that shedding can be efficiently measured after expression of an exogenous substrate by retroviral gene transfer. MCF-7 and MDA-231 cells were infected with retrovirus in which mouse L-selectin was co-expressed with the puromycin selection gene (pBM-L-selectin-IRES-PURO). After infection and selection with puromycin, nearly 100% of the cells expressed

cell surface L-selectin as determined by flow cytometry using a PE-conjugated anti-mouse L-selectin antibody (Figure 4). After stimulation with the phorbol ester (PMA) which is a known activator of TACE-mediated ectodomain shedding, L-selectin was rapidly shed from the cell surface and shedding could be efficiently blocked by pre-incubation with the metalloproteinase inhibitor GM6001 (Figure 4). This experiment confirmed that an exogenous substrate could be efficiently expressed in human breast cancer cells, and that these cell lines contain an active ectodomain sheddase machinery whose activity can be efficiently monitored by flow cytometry.

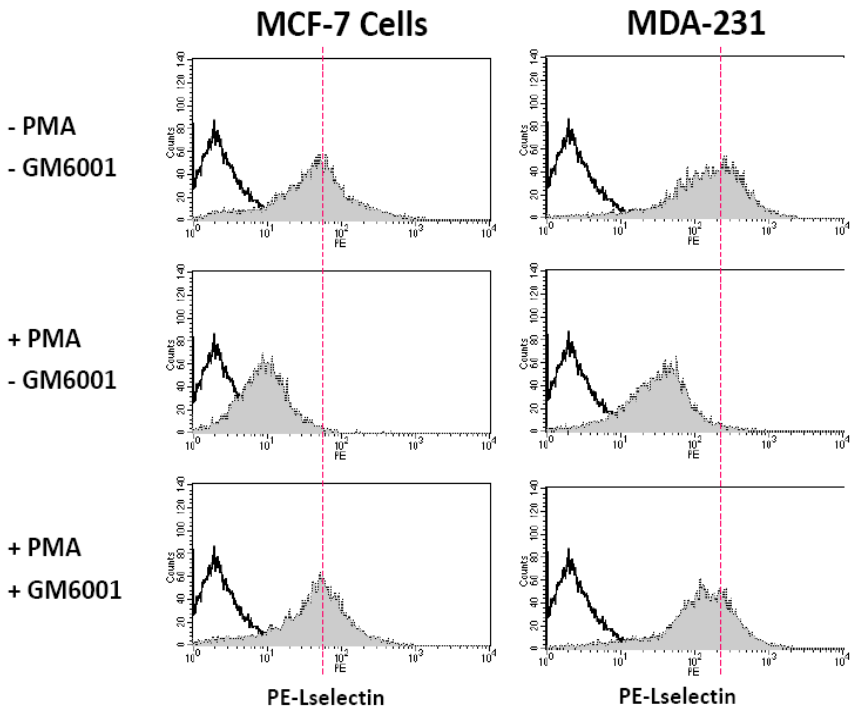


Figure 4: Ectodomain shedding in human breast cancer cells. The breast cancer cell lines MCF-7 and MDA-231 were infected with retrovirus derived from the pBM-Lselectin-IRES-PURO vector and selected with puromycin to obtain selected populations of cells expressing mouse Lselectin. Cell surface levels of L-selectin were monitored by flow cytometry using PE-conjugated antibodies against mouse L-selectin. After stimulation with PMA, L-selectin levels on the cell surface were rapidly down regulated, and shedding could be efficiently blocked with the metalloproteinase inhibitor GM6001.

Generation of a Retroviral Vector for N-terminal Epitope Tagging

A variety of strategies were attempted to generate a retroviral vector that would allow for the expression of N-terminal epitope-tagged proteins that could be subsequently used for the screening of N-terminal tagged cDNA libraries. The first tested vector (pBM-3XTag-IRES-PURO) was designed with a number of features (Figure 5A). The acetylcholine receptor signal sequence was cloned in frame with three tandem epitope tags; the HA epitope tag for which reagents are readily available for immunoprecipitation and Western blotting, a 6-His tag which can be used for Ni-column affinity purification, and a streptavidin binding domain tag that shows very high affinity for streptavidin. These tags were designed to be in-frame with genes cloned between two asymmetric SfiI sites such that library clones could be directionally cloned into the vector backbone. As a test substrate we chose L-selectin as we had previously shown that it is

efficiently expressed after retroviral infection, and can be easily detected after expression by either flow cytometry or by Western blotting when the HA-tag was placed at the C-terminus. L-selectin cDNA was amplified by PCR, digested with SfiI, and subcloned into SfiI linearized pBM-3XTag-IRES-PURO vector. 8 random resulting colonies were analyzed and we found that 8/8 recombinant clones contained the correct L-selectin insert confirming that the asymmetric SfiI cloning strategy led to very efficient generation of recombinant clones (Figure 5B). Retrovirus derived from the resulting vector was used to infect NIH3T3 cells as expression of L-selectin was determined by flow cytometry (Figure 5C). Unlike with untagged or C-terminal tagged L-selectin, we found that L-selectin was not expressed at the cell surface. We therefore used Western blotting to determine if L-selectin was being expressed at all, and using antibodies to either the HA- or the 6His-epitope tag we were able to show high level expression of L-selectin. This confirms that the expression construct was in fact expressing L-selectin, but the expressed protein was not being properly processed to the cell surface (Figure 5D).

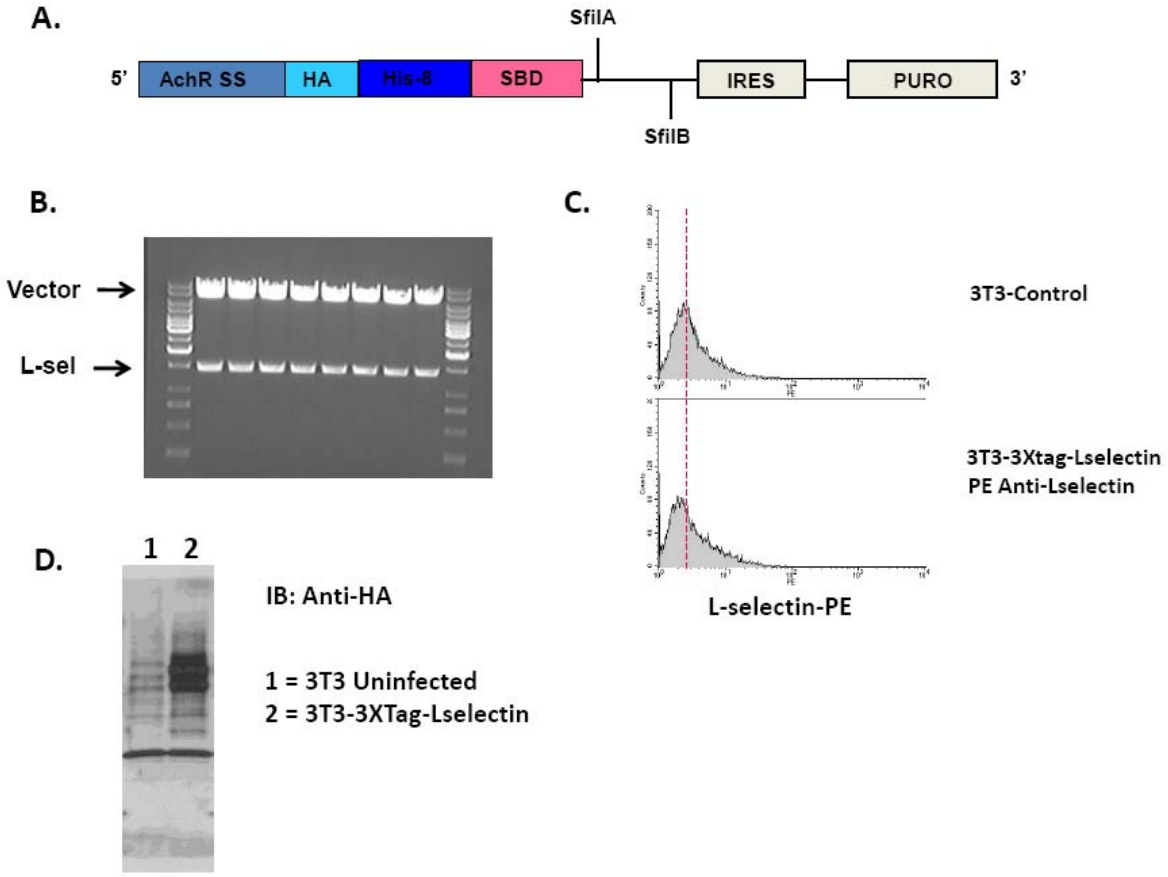


Figure 5: Characterization of the pBM-3XTag retroviral vector. A. Schematic of the pBM-3XTag vector. B. Cloning efficiency after amplification and subcloning of L-selectin into the pBM-3XTag vector. C. Lack of expression of L-selectin at the cell surface in NIH3T3 cells after infection with Lselectin3Xtag retovirus. D. Western blot confirming intracellular expression of N-terminal tagged L-selectin.

Optimization of the N-terminal HA-tagged retroviral vector

To express breast cancer-associated proteins containing an extracellular HA-epitope tag that can be detected by Western blotting or flow cytometry, two retroviral vectors were constructed. First we tested an N-terminal HA-tagged vector for protein expression in

mammalian cells. Secondly, for the type II membrane protein expression, we also tested a C-terminal HA-tagged vector. We chose a retroviral vector system for exogenous gene expression for the following reasons: 1) it provides stable and efficient transduction of the primary cell, 2) the method of the gene transduction is non-toxic, and 3) the gene transduction, gene expression and selection is rapid. In previous experiments, we tested the expression of L-selectin in our retroviral vector with both N-terminal and C-terminal HA tags (Figure 6, clone L-selectin-HA and HA-L-selectin) in mammalian cells. Both N-terminal and C-terminal-tagged target proteins could be detected in total cell lysates by Western blot (Figure 5D and Figure 7, Lane 4). However, only the C-terminal tagged L-selectin protein could be detected on the cell surface by flow cytometry and cell sorting assay. This could be due to protein misfolding caused by the N-terminal tag epitope, or the secreted protein's signal sequence was covered by the epitope tag. To test the latter hypothesis, we made 3 constructs (Figure 6, HA-Lselectin.1, HA-Lselectin.2 and HA-Lselectin.3) in which the HA epitope tag was placed at increasing distances from the N-terminus using PCR and DNA fragment synthesis from Gene Script, Inc. The cDNA encoding these 3 HA-L-selectin hybrid genes were cloned into the pBM-HA-IRES-PURO retroviral vector. The PCR-generated fragments were verified by DNA sequencing ensure no mutations were present.

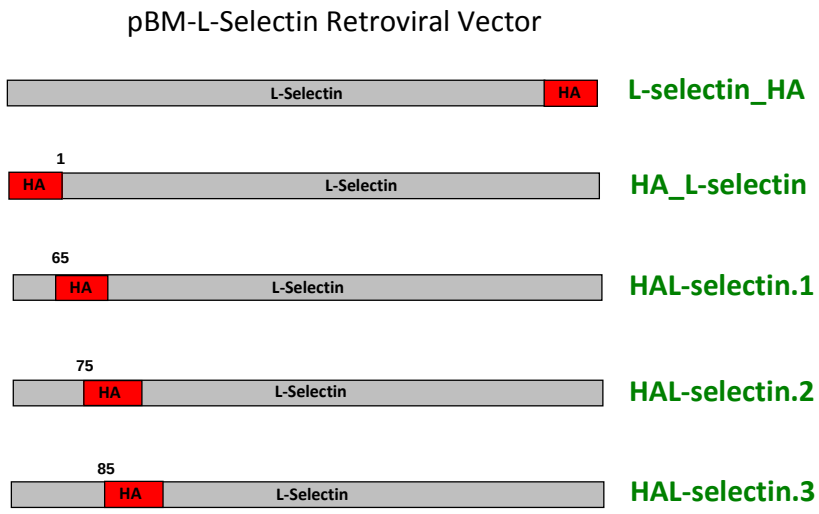


Figure 6: Construction of HAL-selectin.1, HAL-selectin.2 and HAL-selectin.3 clones. cDNAs encoding full length HA-Lselectin were cloned into pBM-IRES-PURO retroviral vector. HA epitope tag was individually tagged into L-selectin amino acid 65, 75 and 85 N-terminal downstream.

In order to generate high titers of retrovirus for infection, all three clones were transiently transfected to human 293T cells to produce retrovirus. Twenty four hours before retrovirus infection, NIH-3T3 mouse cells were plated in 100mm plates at a density of 500,000 cells/plate in DMEM/10% FBS medium. For infection, the growth medium from the NIH-3T3 culture was removed and 5 mL retrovirus supernatant from 293T cells was added. The NIH-3T3 cells were exposed to retrovirus for 12 hours, and then the virus-containing supernatant was replaced with fresh DMEM/FBS growth medium with puromycin (1.5 ug/mL). The cells were cultured for 3 days under puromycin selection so that only cells transfected with the vector containing the HA-L-selectin hybrid genes would grow.

The retrovirus-infected NIH-3T3 cells were harvested by centrifugation. Protein lysates were prepared as follows: 1) RIPA buffer(3) was added to a concentration of 5 x 10^5 cells/ ml, 2) The cells were incubated on ice for 10 minutes, 3) the cell pellets were sonicated 4 times, 4) and then the cell lysate was centrifuged at 20,000 x g for 10 minutes to harvest the supernatent. Lysates were separated on a NuPage 4-12% Tris-Bis Gel (Invitrogen, Carlsbad, CA) and then the proteins were transferred to Hybond-c membrane (Invitrogen, Carlsbad, CA) for Western blotting.

Western blot of retrovirus infected NIH3T3 cell lystae

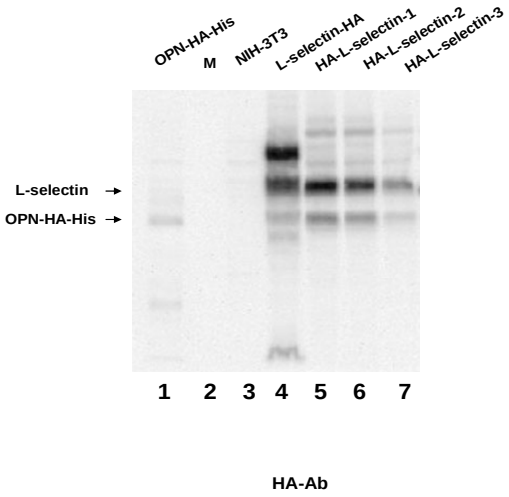
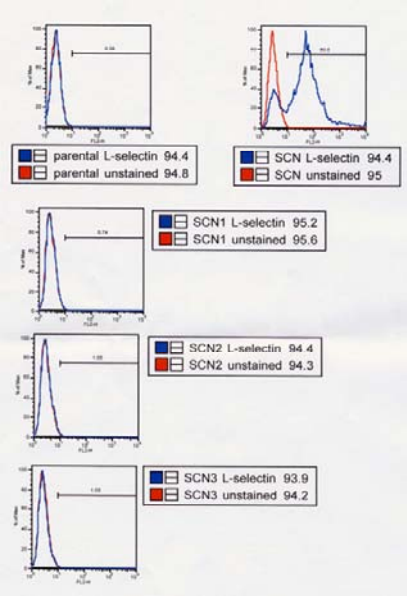


Figure 7: Analysis of L-selectin with an HA epitope tag-infected NIH3T3 cell lysate. The NIH3T3 cells were lysed with RIPA buffer at a density of 50,000 cells/ml. 10 uL of lysate were used for Western blotting with an anti-HA antibody.

Analysis of NIH-3T3 cell surface L-selectin expression by flow cytometry and cell sorting

Retrovirus-infected cells were prepared for analysis of HA-L-selectin protein cell surface expression by harvesting from culture plates by adding 0.05% Trypsin/EDTA (Invitrogen, Carlsbad, CA) for 5 minutes, and then keeping on ice prior to antibody staining. The FACS stain was set up immediately as follows: 1) 200 µL cells (about 200,000 cells) were incubated in a 3X dilution of rat anti-mouse L-selectin antibody (BD Pharmingen, San Diego, CA), 2) cells were stained for 15 minutes in the dark at 4°C, 3) cells were pelleted by centrifugation (1000 g, 5 min.), 4) the supernatant was decanted and the pellet was washed with 200 µL FACS buffer, 5) the cell pellet was resuspended in 2% Paraformaldehyde/PBS buffer after the last wash. The stained and fixed cells were stored in the dark before analyzing with a FACScan flow cytometer (BD Biosciences, San Jose, CA).

FACScan



Antibody:
PE Rat anti-mouse CD62L (MEL-14, L-selectin)
BD Pharmingen Cat. #: 553151

(1 ug per 200,000 cells)

1. **NIH3T3 Parental**
2. **L-selectin-HA**
3. **HA-L-selectin.1**
4. **HA-L-selectin.2**
5. **HA-L-selectin.3**

Figure 8: FACS assay of pBM-HAL-selectin.1, pBM-HAL-selectin.2 and pBM-HAL-selectin.3 infected NIH3T3 cells. NIH3T3 cells were selected with puromycin for 72 h and cell surface levels of each clone were determined by FACS.

As explained previously, we decided to test whether cloning the HA epitope tag various distances downstream from the N-terminal L-selectin secreted signal peptide (amino acid 1-38) would allow cell surface expression of the protein. The results are shown in Figure 4 where only the C-terminally tagged L-selectin-HA clone gives a peak after staining with an anti-L-selectin antibody. However, none of the 3 cDNA encoding full-length L-selectin clones with the HA tag at the N-terminal end were detected by flow cytometry and cell sorting assay after staining with the anti-L-selectin antibody, even as far as 47 amino acid residues from the signal sequence (for clone HA-L-selectin.3).

Construction of C-terminal HA-tagged Retroviral Clones

Since the N-terminal tagged clones could not be detected on the cell surface by FACS, we made a construct in which the HA epitope tag is fused to the C-terminus of the L-selectin gene (pBM-L-selectin-HA-IRES-Puro) and Osteopontin gene (pBM-OPN-HA-His-IRES-Puro). The osteopontin cDNA template was purchased from Harvard Proteomic Institute.

The primers for PCR amplification are upstream forward primer:

5’-GCCGGATCCGCCACCATGAGAATTGCAGTG-3’

and downstream reverse primer:

5’-GATGCGGCCGCCTACTAATGGTGATGATGGTGGTGATG-3’.

The PCR-generated fragments were verified by DNA sequencing to make sure there were no mutations during amplification. The products were then restriction enzyme-digested with BamHI and Not I, and cloned into pBM-HA-IRES-PURO BamHI and NotI sites. (Figure 9)

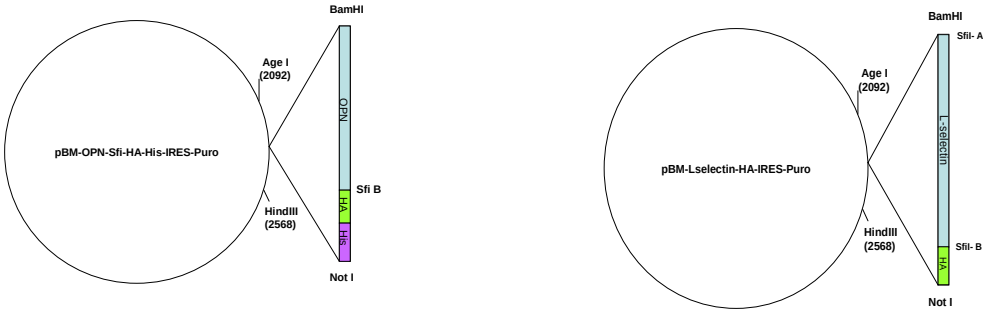
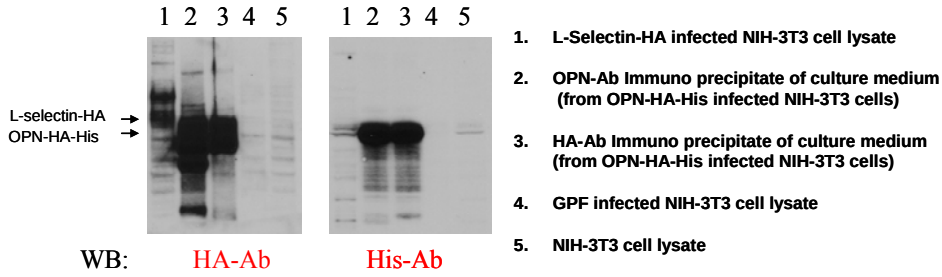


Figure 9: Plasmid maps of pBM-osteopontin-HA-IRES-PURO and pBM-L-selectin-HA-IRES-PURO retroviral clones.

Retrovirus production, cell culture and cell lysates were prepared for the C-terminus tagged clones as described previously. For protein analysis, cells were lysed in 50mM Tris-HCl, pH7.4, 250 mM NaCl, 0.5% NP-40, 10% glycerol, 5mM EDTA, 50mM NaF, 10 mM PMSF, 5 ug/mL leupeptin and aprotinin. Anti-HA antibody (Roche, Indianapolis, IN) was used for Western blotting. The results show L-selectin-HA and Osteopontin-HA-His fused proteins could be detected in the cytoplasm (Figure 10A, left panel). Anti-HA antibody was also used to immunoprecipitate the OPN-HA-His fusion protein from the cell growth medium and then used to detect the OPN-HA-His (Figure 10A, right panel) by Western blot. This suggests the osteopontin-HA-His fusion protein was successfully secreted out of the cell.

A.



B.

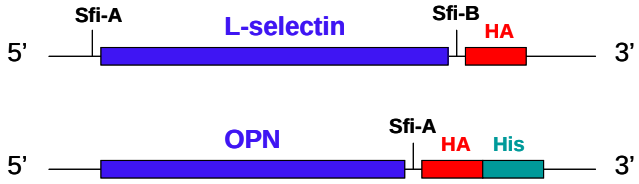


Figure 10: A. Western blotting of cell lysates from pBM-Lselectin-HA and pBM-OPN-HAHis infected NIH3T3 cells. B. cDNA encoding full length L-selectin and Osteopontin were cloned into a pBM-HA-IRES-PURO retroviral vector.

FACS assay was performed for the C-terminus tagged clones as described in the previous section. The infected cells were harvested by trypsinization and washed in PBS buffer. The results shown in Figure 11, indicate that the transgene protein can also be successfully detected by flow cytometry and Western blotting assays when the HA tag is fused to C-terminus of the transgene. Expression of the tagged proteins continues even after 7 passages in culture (data not shown).

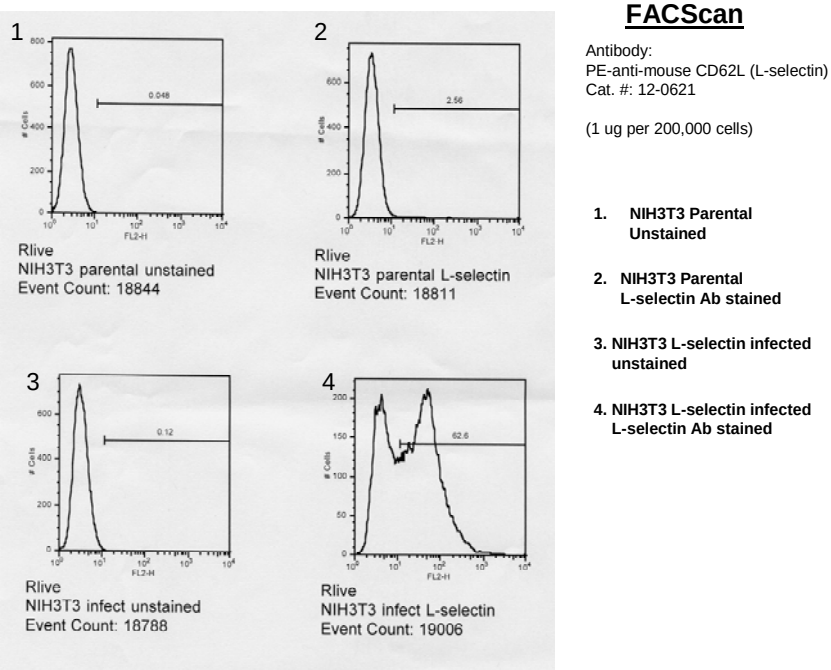


Figure 11: FACS analysis of NIH3T3 cell after infection with pBM retroviral vector containing L-selectin . Cells were selected with puromycin for 72 h before harvesting. Panel1: parental NIH3T3 cells without staining. Panel 2: parental NIH3T3 cells with anti-L-selectin antibody staining. Panel 3: parental infected NIH3T3 cells with pBM retroviral vector containing L-selectin gene without staining. Panel 4: parental infected NIH3T3 cells with pBM retroviral vector containing L-selectin gene with anti L-selectin antibody staining.

Mass spectrometry detection of C-terminus tagged clones

In order to further verify that our tagged proteins were being expressed at sufficient abundance in our cell expression system, we prepared samples for mass spectrometry analysis. Anti-HA antibody coupled to protein A agarose beads were used to immunoprecipitate the L-selectin-HA protein from a cell lysate of retrovirally-infected NIH3T3 cells. The agarose beads were washed with PBS 3 times and then the captured protein was eluted by adding 5% acetic acid (1). The eluted protein was then prepared for mass spectrometry analysis as follows: 1) the protein was denatured in 25 mM Tris pH8.0, 6 M Urea and 10 mM TCEP, 2) the protein was alkylated with 10 mM iodoacetamide, 3) samples were diluted into 25 mM Tris pH 8.0, 0.6 M Urea, 1 mM TCEP 4) and then digested with Trypsin Gold (Promega, Madison, WI) at 1:50 trypsin:protein ratio at 37° overnight. After trypsin digestion, the peptides were desalted and buffer exchanged by binding to a C18 column and then eluting in a solution of 80% acetonitrile, 0.1% formic acid. The samples were then dried in a Speedvac concentrator and resuspended in a solution of 2% acetonitrile, 0.1% formic acid at final concentration 0.3 mg/mL.

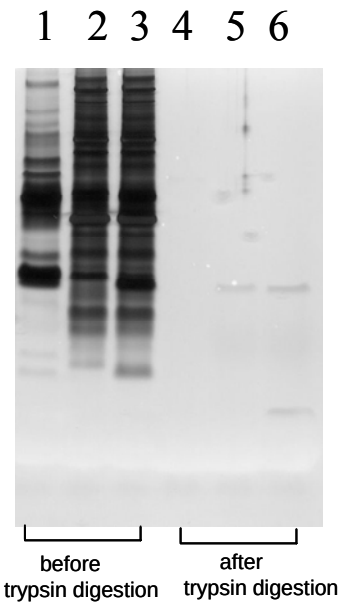


Figure 12: Confirmation of trypsin digestion of HA immunoprecipitated proteins before mass spectrometry analysis. Lane1 and 4: anti-HA immunoprecipitation (IP) samples from NIH-3T3 medium. Lane 2 and 5: IP samples from pBM-IRES-osteopontin-His-HA retrovirus infected NIH-3T3 cell lysate. Lane 3 and 6: IP samples from pBM-IRES-L-selectin-HA retrovirus infected NIH-3T3 cell lysate.

PART B

Strategies for the Synthesis of Retroviral cDNA Libraries

Two strategies were attempted to synthesize a retroviral cDNA library from human breast cancer cells. In the first (Figure 13), pooled mRNA from MDA-231 and MCF7 human breast cancer cell lines was used as a template for first stand cDNA synthesis using random hexamer primers. A random hexamer primer strategy was chosen, as a cDNA library without the endogenous gene stop codons was needed so that the resulting clones would be expressed in-frame with the C-terminal HA tag of the pBM-LIB-HA-IRES-PURO vector. As a result, oligo-dT priming of the C-terminal tagged library could not be done. After first strand synthesis, dC-tailing was done using reverse transcriptase to generate a binding site for the SfiA site-containing SMART primer (Clontech, Figure 13A). The SMART-primer is ultimately used to generate the dsDNA clone through primer extension (Figure 13B) to yield a cDNA clone with a 5'-SfiIA site corresponding to the upstream cloning site of the pBM-LIB-HA-IRES-PURO retroviral vector. SfiIB adapters were then ligated to the resulting cDNA library such that after digestion with SfiI, the resulting cDNA library clones contained asymmetric SfiI sites compatible with the pBM-LIB-HA-IRES-PURO vector. Despite successful amplification of cDNA, ligation into the pBM-LIB-HA-IRES-PURO vector proved problematic and only appeared to occur at very low frequencies (not shown). To circumvent the potential limitations of adapter ligation, a second attempt was made by using a random hexamer primer into which the downstream SfiI site had been incorporated, circumventing the need for adapter ligation (Figure 13). Despite this change, only clones representing a non-specific recombination of the pBM-LIB-HA-IRES-PURO vector were obtained (data not shown).

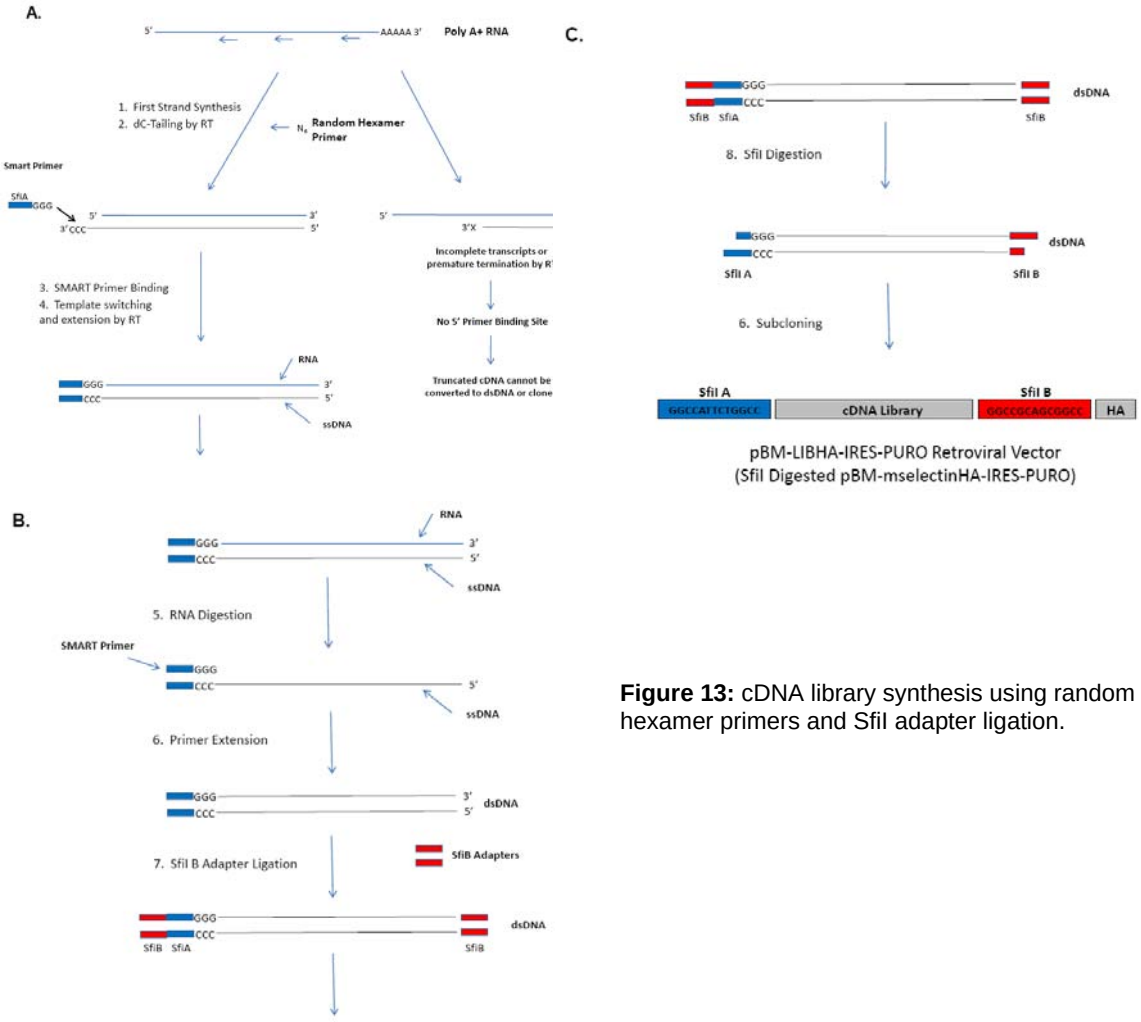


Figure 13: cDNA library synthesis using random hexamer primers and SfiI adapter ligation.

Generation of HA-tagged breast cancer-associated and kinase libraries

The challenges of generating a tagged-ORF library for this project are: 1) The tagged sequence needs to be in-frame with the ORF; 2) the coding region must be long enough to have a domain that is secreted or translocated to a subcellular compartment; 3) the N-terminal tag must not affect protein trafficking to the membrane or subcellular compartment; 4) to create a C-terminal tagged ORF, the tagged sequence joining the open reading frame must not contain the endogenous stop codon. As described in **II. Body, Part A**, we have determined that moving the tagged sequence within the coding sequence of L-selectin blocks the expressed protein from being translocated to the cell surface. We had contracted a private company to generate a C-terminal HA-tagged breast cancer library from MCF7 and MDA-MB-231 cell lines. First the mRNA was isolated, reverse transcribed with poly-d(T) primer and subsequently PCR-amplified using primers containing random hexamer-SfiI to generate double-stranded cDNA. The SfiI-cDNA was cloned into our retroviral vector as described in **II. Body, Part B**. However, the company failed to generate clones with a significant amount of genes to represent the whole

cDNA library. As a result, we did not have the opportunity to proceed to mass spectrometry analysis of proteins shed or secreted from a custom cDNA breast cancer library.

As an alternative to this custom cDNA breast cancer library, we acquired two human cDNA open reading frame (ORF) libraries from Harvard Institute of Proteomics. One of the collections, a human breast cancer associated ORF library (Breast Cancer 1000 collection) (5), includes 2,180 genes significantly changed in breast cancer cells. Because many breast cancer cases involve the elevation of kinase activity of the EGFR/Her2 signaling pathway, and because many kinases are found to be activated and associated with cancer initiation, acceleration, and progression, we also acquired a "human kinase ORF library" (4), which includes 697 kinase genes. Both cDNA clone libraries can be subcloned into our retroviral vector by PCR-amplifying the ORF region.

To clone these ORFs into our retroviral vector, we first designed primers with SfiI sites at both 5' and 3'- ORF flanking regions. The PCR-amplified ORFs with SfiI sites were subcloned into the retroviral vector to make HA-tagged fusion proteins in three categories. Figure 14 illustrates the experimental procedure to generate HA-tagged ORF library.

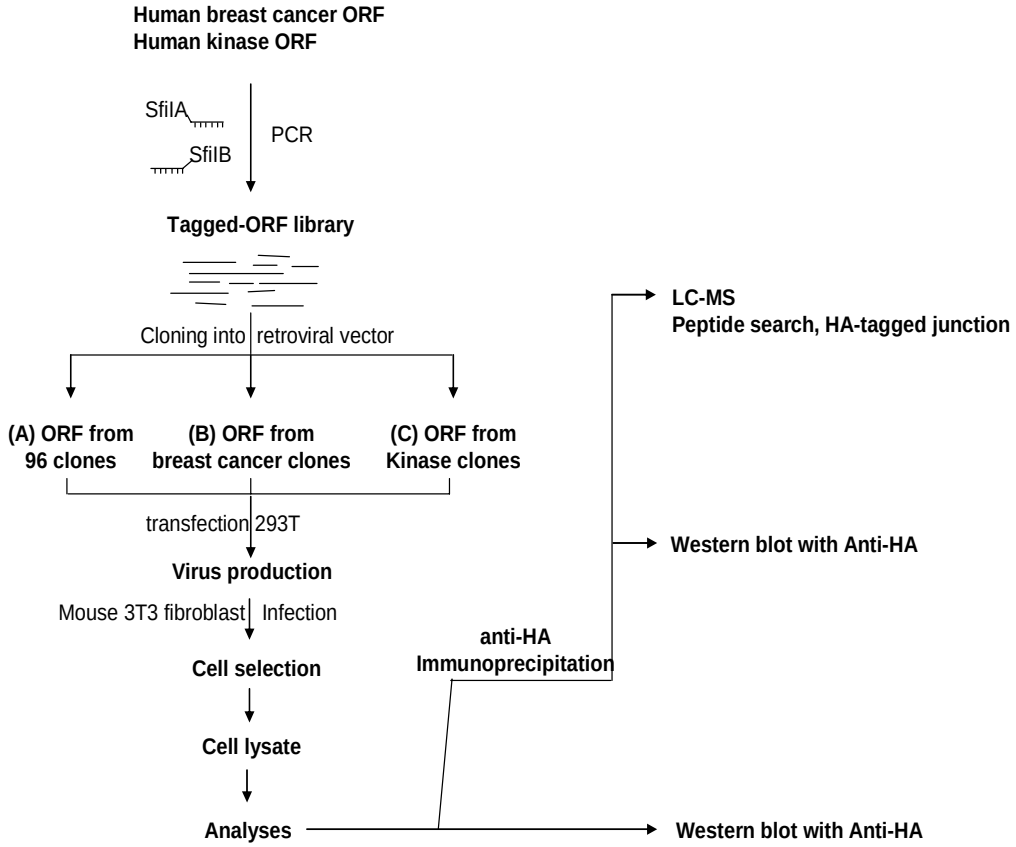


Figure 14: Construction of HA-tagged human breast cancer-associated and kinase open reading frame library.

To prove the primers we designed can be used to amplify the library, we picked one 96-well plate from the Breast Cancer 1000 library for template and used both SfiA (pDNR-Fwd-Sfi: 5'-ATCCGGCCATTCTGGCCTATACGAAGTTATCAGTCGACACCATG-3') and SfiB (pDNR-Rev-Sfi: 5'-TAGGCCGCTGCGGCCGCGCCAAACGAATGGTCTAGAAAGCTTCCCAA-3') oligo

primers to do individual PCR reactions (Figure 15). The data shows both oligos can be used to amplify the library. We also picked 4 clones to PCR amplify and did the Sfi-I digestion, followed by ligation into the pBM-IRES-Puro retroviral vector. Mini-prep data shows both retroviral vector and oligo primers can be used to transfer the Breast Cancer 1000 and Human Kinase collection into our retrovirus expression vector (Figure 16).

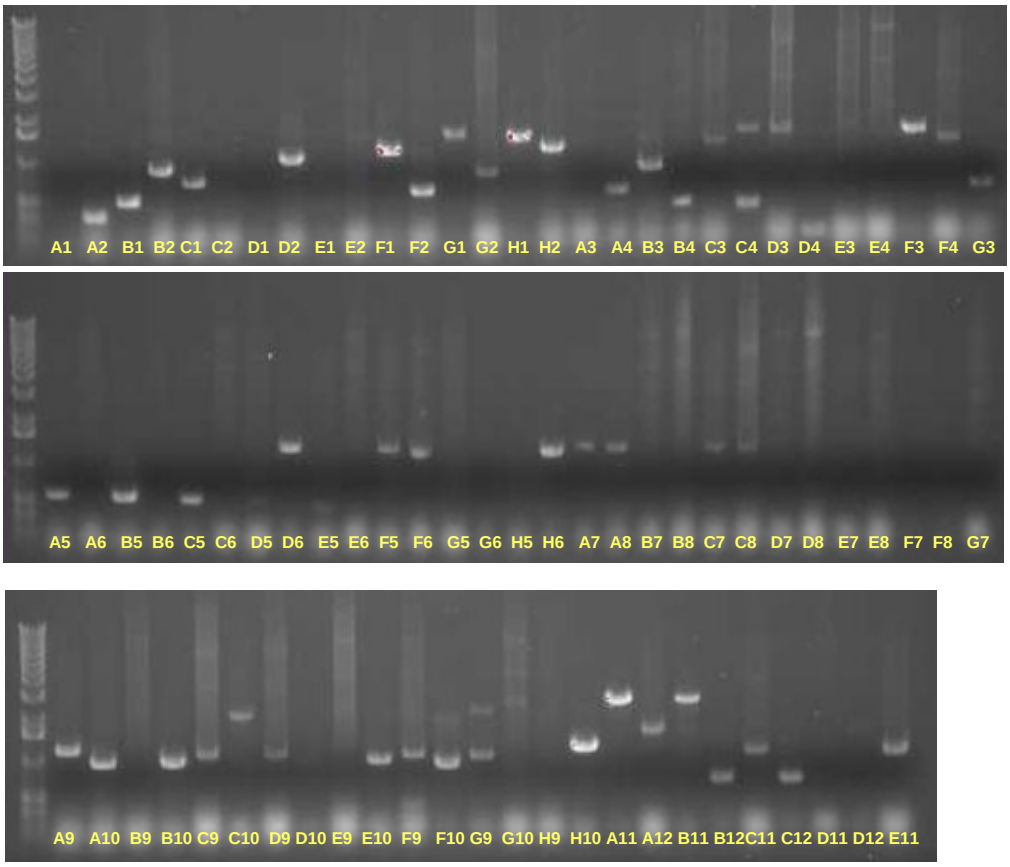


Figure 15: PCR reaction to amplify clones (plate number: BGH1403) from the Breast Cancer 1000 collection from Harvard Proteomic Institute. The footnote below each lane indicates the clone position from the original 96-well plate.

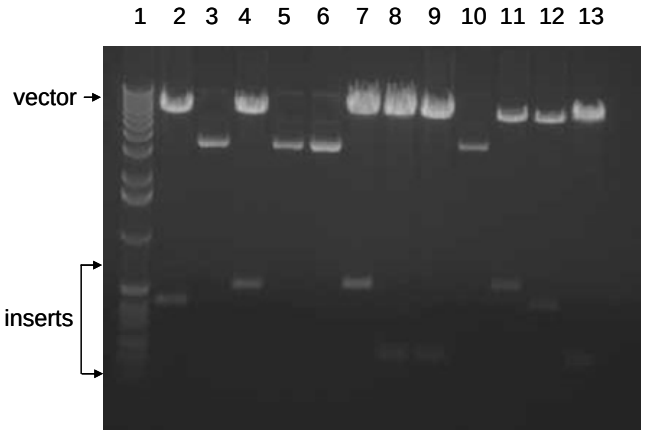


Figure 16: Mini-prep DNA from pBM-IRES-BGH1403 inserts-Puro clones. Clones were randomly picked from the ligation master plate and digested with SfiI restriction enzyme.

III. Key Research Accomplishments

A number of key milestones were reached that will enable alternative strategies to pursue our original goal – the identification of novel serum biomarkers through proteomic screens of retroviral libraries. Toward this end we have accomplished the following:

1) We have shown retroviral infection is an efficient mechanism for gene transfer in human breast cancer cell lines leading to stable and high level protein expression.
2) We have shown that human breast cancer cells have an active ectodomain sheddase machinery and are able to release proteins from the cell surface through protease-mediated ectodomain shedding.
3) We have generated a retroviral vector that allows for efficient C-terminal tagging of proteins with efficient selection of transduced clones through puromycin selection.
4) We have shown that C-terminal HA-tagged proteins can be efficiently immunoprecipitated and purified using high-affinity anti-HA antibodies.
5) We have shown that for some proteins (L-selectin) placement of an epitope tag at the N-terminus appears to disrupt endogenous signal sequence function and therefore normal protein trafficking to the cell surface.
6) Initial attempts at generating cDNA libraries by traditional random hexamer primed methods was unsuccessful, yet we have identified a number of additional strategies in which pre-made libraries can be cloned into retroviral vectors for subsequent expression and screening.
7) We have shown the C-terminal HA-tagged vector can be used to transfer the Breast Cancer 1000 collection genes and could be used for generation of retrovirus in mammalian cells for future screening assay.

IV. Conclusion

Although we encountered numerous technical difficulties, a number of important milestones were reached as outlined above. Importantly, we still feel that the basic hypothesis or premise of this project is sound – that identification of secreted and/or shed proteins through proteomic-based genetic screens may be an effective strategy for the identification of novel soluble biomarkers for human breast cancer that have been difficult to reliably identify by other means. Through this project we were able to generate and test a number of retroviral expression vectors, and have shown that expression of C-terminal tagged proteins is feasible while N-

terminal tagging proved problematic. In addition, future strategies to subclone pre-made cDNA libraries into C-terminal tagging vectors will likely yield more promising libraries for proteomic screens.

V. References

1. J. R. Whiteaker et al., J Proteome Res. 6, 4535 (2007)

2. M. J. Duffy et al., Clin Cancer Res. 15, 1140 (2009).

3. K. J. Garton et al., J Biol Chem. 276, 37993-38001 (2001).

4. A. E. Witt et al., J Proteome Res. 5, 599 (2006). (PMCID: PMC2522320)